Performance Analysis of Sparse Recovery Based on Constrained Minimal Singular Values

Gongguo Tang, *Member, IEEE* and Arye Nehorai, *Fellow, IEEE*

Abstract—The stability of sparse signal reconstruction is investigated in this paper. We design efficient algorithms to verify the sufficient condition for unique ℓ_1 sparse recovery. One of our algorithm produces comparable results with the state-of-the-art technique and performs orders of magnitude faster. We show that the ℓ_1-constrained minimal singular value (ℓ_1-CMSV) of the measurement matrix determines, in a very concise manner, the recovery performance of ℓ_1-based algorithms such as the Basis Pursuit, the Dantzig selector, and the LASSO estimator. Compared with performance analysis involving the Restricted Isometry Constant, the arguments in this paper are much less complicated and provide more intuition on the stability of sparse signal recovery. We show also that, with high probability, the subgaussian ensemble generates measurement matrices with ℓ_1-CMSVs bounded away from zero, as long as the number of measurements is relatively large. To compute the ℓ_1-CMSV and its lower bound, we design two algorithms based on the interior point algorithm and the semi-definite relaxation.

Index Terms—ℓ_1-constrained minimal singular value, Basis Pursuit, Dantzig selector, interior point algorithm, LASSO estimator, restricted isometry property, sparse signal reconstruction, semidefinite relaxation, verifiable sufficient condition

I. INTRODUCTION

Sparse signal reconstruction aims at recovering a sparse signal $x \in \mathbb{R}^n$ from observations of the following model:

$$y = Ax + w, \qquad (1)$$

where $A \in \mathbb{R}^{m \times n}$ is the measurement or sensing matrix, y is the measurement vector, and $w \in \mathbb{R}^m$ is the noise vector. The sparsity level k of x is defined as the number of non-zero components of x. The measurement system is underdetermined because the number of measurements m is much smaller than the signal dimension n. However, when the sparsity level k is also small, it is possible to recover x from y in a stable manner. Reconstruction of a sparse signal from linear measurements appears in many signal processing branches, such as compressive sensing [1]–[3], sparse linear regression [4], source localization [5], [6], sparse approximation, and signal denoising [7]. Model (1) is applicable to many practical areas such as DNA microarrays [8]–[11], radar imaging [12]–[14], cognitive radio [15], and sensor arrays [5], [6], to name a few.

This paper is motivated by the following considerations. When we are given a sensing or measurement system (1),

we usually want to know the performance of the system before using it, or at least to know whether it works in the ideal setting. This involves deriving a verifiable sufficient condition for unique recovery and a computationally amenable performance measure. Furthermore, in signal processing and the applications mentioned in the previous paragraph, we usually have the freedom to design the sensing matrix; that is, we can choose the best from a collection of sensing matrices. For example, in radar imaging and sensor array applications, sensing matrix design is connected with waveform design and array configuration design. The ℓ_1-CMSV of this paper has already been used to optimally design orthogonal frequency division multiplexing (OFDM) radar signals for detecting a moving target in the presence of multipath reflections [16]. We hope the practitioners in similar fields will also find the results in this paper useful.

The contribution of the work is fourfold. First, we design two algorithms to verify the sufficient condition that a sparse signal can be uniquely recovered using ℓ_1 minimization in the noise-free setting. By solving multiple linear programs efficiently, one of the algorithms produces comparable results with the state-of-the-art verification algorithms and performs orders of magnitude faster. Second, we derive concise bounds on the ℓ_2 norm of estimation error for the Basis Pursuit, the Dantzig selector, and the LASSO estimator in terms of the ℓ_1-CMSV. As the third contribution, we demonstrate that if the number of measurements m is reasonably large, subgaussian random matrices have ℓ_1-CMSVs bounded away from zero, with high probability. This implies that at least for subgaussian random matrices, the ℓ_1-CMSV is as good as the restricted isometry constant. Last but not least, we develop algorithms to compute the ℓ_1-CMSV for an arbitrary sensing matrix and compare their performance. These algorithms are by no means the most efficient ones. However, once we shift from an optimization problem with a discrete nature (*e.g.*, the restricted isometry constant) to a continuous one, there are many optimization tools available and more efficient algorithms can be designed.

Many quantities and properties on the sensing matrix A have been proposed to guarantee a stable or unique signal reconstruction, most notably the Restricted Isometry Constant (RIC) [1], [17] and the Null Space Property (NSP) [18]. The RIC provides a unified framework to deal with sparse signal recovery and has very nice geometrical explanations. However, it is known to be very difficulty to compute. Therefore, computable bounds on quantities involved in the RIC and the NSP are computed using, for example, semi-definite programming relaxation [19], [20], and linear programming

relaxation [21]. To the best of the authors' knowledge, the algorithms of [20] and [21] in verifying the sufficient condition of unique ℓ_1 recovery represent the state-of-the-art technique in this direction. In our paper, instead of computing a quantity (α_k in [21]) for various sparsity levels k and checking if it is less than $1/2$, we directly seek the critical sparsity level below which unique recovery is guaranteed. We compare our verification algorithms with those in [20] and [21] using numerical simulations. One of our algorithms performs orders of magnitude faster, consumes much less memory, and produces comparable results.

The paper is organized as follows. In Section II, we present the measurement model, three convex relaxation algorithms, and the RIC. Section III is devoted to deriving sufficient conditions for unique ℓ_1 recovery and deriving bounds on the recovery errors of several convex relaxation algorithms. In Section IV, we show that the majority of realizations of the subgaussian measurement ensemble have good ℓ_1-CMSVs. In Section V, we design algorithms to verify unique recovery, and compute the ℓ_1-CMSV and its lower bound. We compare the algorithms' performance in Section VI. Section VII summaries our conclusions.

II. The Measurement Model, Reconstruction Algorithms, and Restricted Isometry Property

A. The Measurement Model

The following measurement model is used throughout the paper. Suppose we have a sparse signal $\boldsymbol{x} \in \mathbb{R}^n$, *i.e.*, $\boldsymbol{x}$ has only a few non-zero components. The sparsity level k of $\boldsymbol{x}$ is defined as the number of non-zero elements of $\boldsymbol{x}$, or the ℓ_0 "norm" of $\boldsymbol{x}$: $k = \|\boldsymbol{x}\|_0$. We call a vector $k-$sparse if its sparsity level $\|\boldsymbol{x}\|_0 \leq k$. For ease of presentation, we restrict ourselves to exactly sparse signals in this paper and leave approximately sparse signals to future work.

We observe $\boldsymbol{y} \in \mathbb{R}^m$ through the following linear model:

$$\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{w}, \tag{2}$$

where $A \in \mathbb{R}^{m \times n}$ is the sensing/measurement matrix and $\boldsymbol{w} \in \mathbb{R}^m$ is the noise/disturbance vector.

In this paper, we focus on three most renown recovery algorithms based on convex relaxation: the Basis Pursuit (BP) [22], the Dantzig Selector (DS) [23], and the LASSO estimator [24]:

$$\text{BP: } \min_{\boldsymbol{z} \in \mathbb{R}^n} \|\boldsymbol{z}\|_1 \text{ s.t. } \|\boldsymbol{y} - A\boldsymbol{z}\|_2 \leq \epsilon \tag{3}$$

$$\text{DS: } \min_{\boldsymbol{z} \in \mathbb{R}^n} \|\boldsymbol{z}\|_1 \text{ s.t. } \|A^T(\boldsymbol{y} - A\boldsymbol{z})\|_\infty \leq \lambda_n \sigma \tag{4}$$

$$\text{LASSO: } \min_{\boldsymbol{z} \in \mathbb{R}^n} \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{z}\|_2^2 + \lambda_n \sigma \|\boldsymbol{z}\|_1. \tag{5}$$

Here λ_n is a turning parameter, and ϵ and σ is a measure of the noise level. All these three optimization programs can be implemented efficiently using convex programming or even linear programming.

The performance of the BP, the DS and the LASSO, more specifically the error bounds on the solutions of these algorithms, usually involve the incoherence of the sensing matrix A. Many quantities are proposed to measure the incoherence

of the sensing matrix, for example, the Restricted Isometry Constant (RIC) [3], [17], the Restricted Eigenvalue assumption [25], and the Restricted Correlation assumption [26], among others. However, these quantities are very difficult to compute. For example, the only known technique to exactly compute the RIC is test all its submatrices of certain size.

We will compare our CMSV based bounds with the RIC based bounds. For this purpose, we follow [3], [17] to define the RIC as follows:

Definition 1 *For each integer $k \in \{1, \ldots, n\}$, the restricted isometry constant (RIC) δ_k of a matrix $A \in \mathbb{R}^{m \times n}$ is defined as the smallest $\delta > 0$ such that*

$$1 - \delta \leq \frac{\|A\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} \leq 1 + \delta \tag{6}$$

holds for arbitrary non-zero $k-$sparse signal $\boldsymbol{x}$.

The RIC has very clear geometrical meanings. A matrix A with a small δ_k roughly means that A is nearly an isometry when restricted onto all $k-$sparse vectors.

Now we cite some of the most renown performance results on the BP, the DS, and the LASSO, which are expressed in terms of the RIC. Assume $\boldsymbol{x}$ is a $k-$sparse signal and $\hat{\boldsymbol{x}}$ is its estimate given by any of the three algorithms; then we have the following:

1) BP [17]: Suppose that $\delta_{2k} < \sqrt{2} - 1$ and $\|\boldsymbol{w}\|_2 \leq \epsilon$. The solution to the BP (3) satisfies

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{4\sqrt{1 + \delta_{2k}}}{1 - (1 + \sqrt{2})\delta_{2k}} \cdot \epsilon. \tag{7}$$

2) DS [23]: If the noise $\boldsymbol{w}$ satisfies $\|A^T\boldsymbol{w}\|_\infty < \lambda_n \sigma$, and $\delta_{2k} + \delta_{3k} < 1$. Then, the error signal obeys

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{4\sqrt{k}}{1 - \delta_{2k} - \delta_{3k}} \lambda_n \sigma. \tag{8}$$

3) LASSO [27]: Consider the noise-free case. Under the condition of incoherence design with a sparsity multiplier sequence e_n, the error associated with the LASSO estimator $\hat{\boldsymbol{x}}$ is bounded for sufficiently large n by

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq 17.5 \cdot \lambda_n \sigma \cdot \frac{\sqrt{k_n}}{(\nu_{e_n k_n}^{\min})^2}. \tag{9}$$

Here, the sparsity level $k = k_n$ depends on n. Refer to [27] for more information on incoherence design and multiplier sequence.

We note that in these error bounds, the terms involving the RIC on the right hand sides are quite complicated. We will compare these results with our bounds in Section III, which are much more concise and whose derivations are much less involved.

Although the RIC provides a measure quantifying the goodness of a sensing matrix, as mentioned earlier, its computation poses great challenge. The computation difficulty is compensated by the nice properties of RIC for a large class of random sensing matrices. We cite one general result below [28]:

- Let $A \in \mathbb{R}^{m \times n}$ be a random matrix whose entries are *i.i.d.* samples from any distribution that satisfies the concentration inequality for any $\boldsymbol{x} \in \mathbb{R}^n$ and $0 < \varepsilon < 1$:

$$\mathbb{P}\left(\left|\|A\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2\right| \geq \varepsilon\|\boldsymbol{x}\|_2^2\right) \leq 2e^{-mc_0(\varepsilon)}. \quad (10)$$

Then, for any given $\delta \in (0,1)$, there exist constants $c_1, c_2 > 0$ depending only on δ such that $\delta_k \leq \delta$, with probability not less than $1 - 2e^{-c_2 m}$, as long as

$$m \geq c_1 k \log \frac{n}{k}. \quad (11)$$

We remark that distributions satisfying the concentration inequality (10) include the Gaussian distribution and the Bernoulli distribution. For the ℓ_1-CMSV, we will establish a theorem similar to the one above for the subgaussian ensemble with the same bound on m. The subgaussian ensemble in this paper includes the Gaussian ensemble, the Bernoulli ensemble, as well as the normalized volume measure on various convex symmetric bodies , for example, the unit balls of ℓ_p^n for $2 \leq p \leq \infty$ [29].

III. STABILITY OF CONVEX RELAXATION BASED ON THE ℓ_1-CONSTRAINED MINIMAL SINGULAR VALUE

In this section, we present two approaches to verify the sufficient condition for the uniqueness of ℓ_1-recovery. We also derive bounds on the reconstruction errors for the BP, the DS and the LASSO. Our bounds are given in terms of the ℓ_1-CMSV rather than the RIC of matrix A .

We first introduce a quantity that measures the sparsity, (or, more accurately, the density), of a given vector $\boldsymbol{x}$.

Definition 2 *The ℓ_1-sparsity level of a non-zero vector $\boldsymbol{x} \in \mathbb{R}^n$ is defined as*

$$s(\boldsymbol{x}) = \frac{\|\boldsymbol{x}\|_1^2}{\|\boldsymbol{x}\|_2^2}. \quad (12)$$

The scaling and permutation invariant $s(\boldsymbol{x})$ is indeed a measure of sparsity. To see this, suppose $\|\boldsymbol{x}\|_0 = k$; then the Cauchy-Schwartz inequality implies that

$$s(\boldsymbol{x}) \leq k, \quad (13)$$

and we have equality if and only if the absolute values of all non-zero components of $\boldsymbol{x}$ are equal. Therefore, the more non-zero elements $\boldsymbol{x}$ has and the more evenly the magnitudes of these non-zero elements are distributed, the larger $s_p(\boldsymbol{x})$. In particular, if $\boldsymbol{x}$ has exactly one non-zero element, then $s(\boldsymbol{x}) = 1$; if $\boldsymbol{x}$ has n non-zero elements with the same magnitude, then $s(\boldsymbol{x}) = n$.

We use the ℓ_1-sparsity level as a tool to relax the necessary and sufficient condition for exact ℓ_1 recovery in the noise less setting [30]–[32]. In particular, Zhang showed in [30] that $\boldsymbol{x}$ with $\|\boldsymbol{x}\|_0 = k$ is the unique solution to BP with $\epsilon = 0$:

$$\min_{\boldsymbol{z} \in \mathbb{R}^n} \|\boldsymbol{z}\|_1 \text{ s.t. } A\boldsymbol{x} = A\boldsymbol{z} \quad (14)$$

if and only if

$$\sum_{i \in S} |\boldsymbol{z}_i| < \sum_{i \notin S} |\boldsymbol{z}_i| \quad (15)$$

for any $\boldsymbol{z}$ such that $A\boldsymbol{z} = 0$ and any index set $S \subset \{1, \ldots, n\}$ of size at most k. We are interested in finding k^*, the maximal k such that the necessary and sufficient condition (15) is satisfied.

We note that a sufficient condition for exact ℓ_1 recovery is

$$s(\boldsymbol{z}) > 4k \quad (16)$$

for any $\boldsymbol{z} \in \mathrm{Ker}(A) \stackrel{\text{def}}{=} \{\boldsymbol{z} : A\boldsymbol{z} = 0\}$. This is because the negation of (15):

$$\exists \boldsymbol{z} \in \mathrm{Ker}(A) \text{ and } S \text{ with size at most } k$$
$$\text{such that } \sum_{i \in S} |\boldsymbol{z}_i| \geq \sum_{i \notin S} |\boldsymbol{z}_i|$$

implies

$$\begin{aligned}
\|\boldsymbol{z}\|_1 &\leq 2 \sum_{i \in S} |\boldsymbol{z}_i| \\
&\leq 2\sqrt{k} \sqrt{\sum_{i \in S} |\boldsymbol{z}_i|^2} \\
&\leq 2\sqrt{k} \|\boldsymbol{z}\|_2.
\end{aligned}$$

Therefore, the minimization of $s(\boldsymbol{z})$ over $\mathrm{Ker}(A)$ yields a lower bound on k^*. Unfortunately, this optimization is very difficult. In section V, we present a semidefinite relaxation algorithm to obtain a lower bound on k^*.

Another relaxation approach is to replace the ℓ_2 norm in the definition of the ℓ_1-sparsity level with the ℓ_∞ norm. Note that the negation of (15) also implies that

$$\begin{aligned}
\|\boldsymbol{z}\|_1 &\leq 2 \sum_{i \in S} |\boldsymbol{z}_i| \quad (17) \\
&\leq 2k \|\boldsymbol{z}\|_\infty. \quad (18)
\end{aligned}$$

Therefore, the following optimization problem

$$\min_{\boldsymbol{z}: A\boldsymbol{z} = 0} \frac{1}{2} \frac{\|\boldsymbol{z}\|_1}{\|\boldsymbol{z}\|_\infty} \quad (19)$$

finds a lower bound on the maximal k such that (15) is satisfied. In Section V, we will present a polynomial time algorithm to solve (19). The algorithm solves n linear programs and produces results comparable to the best known results in [21] within a much shorter time.

In the noisy setting, our derivation of the error bounds for the BP, the DS, and the LASSO relies heavily on the fact that the error vectors have small ℓ_1−sparsity levels.

Now we are ready to define the ℓ_1−constrained minimal singular value (CMSV):

Definition 3 *For any $s \in [1, n]$ and matrix $A \in \mathbb{R}^{m \times n}$, define the ℓ_1-constrained minimal singular value (abbreviated as ℓ_1-CMSV) of A by*

$$\rho_s(A) \stackrel{\text{def}}{=} \min_{\boldsymbol{x} \neq 0, \ s(\boldsymbol{x}) \leq s} \frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}. \quad (20)$$

Intuitively, the ℓ_1−CMSV $\rho_s(A)$ measures the invertibility of the operator $A : \mathbb{R}^n \mapsto \mathbb{R}^m$ when restricted onto vectors with ℓ_1-sparsity level not greater than s.

In the following theorem, we present our error bounds in terms of the ℓ_1−CMSV, whose proof is given in Appendix A:

Theorem 1 *Suppose the support of the true signal $\boldsymbol{x}$ is of size k.*

1) *If the noise $\boldsymbol{w}$ is bounded; that is, $\|\boldsymbol{w}\|_2 \leq \epsilon$, then the solution $\hat{\boldsymbol{x}}$ to the BS* (3) *obeys*

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \;\leq\; \frac{2\epsilon}{\rho_{4k}}. \tag{21}$$

2) *If the noise $\boldsymbol{w}$ in the DS* (4) *satisfies $\|A^T\boldsymbol{w}\|_\infty \leq \lambda_n \sigma$, then the solution to* (4) *obeys*

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{4\sqrt{k}}{\rho_{4k}^2}\lambda_n \sigma. \tag{22}$$

3) *If the noise $\boldsymbol{w}$ in the LASSO* (5) *satisfies $\|A^T\boldsymbol{w}\|_\infty \leq \kappa\lambda_n\sigma$ for some $\kappa \in (0,1)$, then the solution $\hat{\boldsymbol{x}}$ to the LASSO estimator* (5) *obeys*

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{1+\kappa}{1-\kappa} \cdot \frac{2\sqrt{k}}{\rho^2_{\frac{4k}{(1-\kappa)^2}}}\lambda_n\sigma. \tag{23}$$

As shown in Appendix A, the procedure of establishing Theorem 1 has two steps:

1) For all three algorithms, show that the error vector $\boldsymbol{h} = \hat{\boldsymbol{x}} - \boldsymbol{x}$ is ℓ_1−sparse: $s(x) \leq s$, where $s = 4k$ for the BP and the DS, and $s = 4k/(1-\kappa)^2$ for the LASSO. This automatically leads to a lower bound $\|A\boldsymbol{h}\|_2 \geq \rho_s\|\boldsymbol{h}\|_2$;
2) Obtain an upper bound on $\|A\boldsymbol{h}\|_2$ or $\|A\boldsymbol{h}\|_2^2$ and invoke Definition 3 of the ℓ_1-CMSV .

The derivation is simpler than those employed for obtaining the RIC based bounds.

When the noise $\boldsymbol{w} \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_m)$, as shown by Candés and Tao in [23], with high probability, $\boldsymbol{w}$ satisfies the orthogonality condition

$$|\boldsymbol{w}^T A_j| \;\leq\; \lambda_n\sigma \quad \text{for all } 1 \leq j \leq n, \tag{24}$$

for $\lambda_n = \sqrt{2\log n}$. More specifically, defining the event

$$E \stackrel{\text{def}}{=} \{\|A^T\boldsymbol{w}\|_\infty \leq \lambda_n\sigma\}, \tag{25}$$

we have

$$\mathbb{P}(E^c) \;\leq\; \frac{2n \cdot (2\pi)^{-1/2} e^{-\lambda_n^2/2}}{\lambda_n}. \tag{26}$$

Therefore, with $\lambda_n = \sqrt{2(1+t)\log n}$, we obtain

$$\mathbb{P}(E) \geq 1 - \left(\sqrt{\pi(1+t)\log n} \cdot n^t\right)^{-1}. \tag{27}$$

As a consequence, the conditions on noise in Theorem 1 for the DS and the LASSO hold with high probability.

Compared with the RIC bounds (7), (8), and (9), our CMSV bounds (21), (22), and (23) are more concise. Of course, if the CMSV $\rho_s(A)$ is not bounded away from zero, these concise bounds would not offer much. We will show in Section IV that, at least for a large class of random matrices, the corresponding ℓ_1-CMSVs are bounded away from zero with high probability if $m \geq c_1 k \log \frac{n}{k}$.

IV. ℓ_1-CONSTRAINED MINIMAL SINGULAR VALUES OF RANDOM MATRICES

This section is devoted to analyzing the property of the ℓ_1-CMSVs for the subgaussian ensemble. We employ a recent estimate on the behavior of empirical processes involving subgaussian random variables [29].

Before we turn to the general empirical process result of [29] developed by the delicate use of the powerful generic chaining idea, we need some notations and definitions. For a scalar random variable X, the Orlicz ψ_2 norm is defined as

$$\|X\|_{\psi_2} = \inf\left\{t > 0 : \mathbb{E}\exp\left(\frac{|X|^2}{t^2}\right) \leq 2\right\}. \tag{28}$$

Markov's inequality immediately gives that X with finite $\|X\|_{\psi_2}$ has subgaussian tail

$$\mathbb{P}(|X| \geq t) \leq 2\exp(-ct^2/\|X\|_{\psi_2}). \tag{29}$$

The converse is also true, *i.e.,* if X has subgaussian tail $\exp(-t^2/K^2)$, then $\|X\|_{\psi_2} \leq cK$.

A random vector $\boldsymbol{X} \in \mathbb{R}^n$ is called *isotropic and subgaussian* if $\mathbb{E}|\langle \boldsymbol{X}, \boldsymbol{u}\rangle|^2 = \|\boldsymbol{u}\|_2^2$ and $\|\langle \boldsymbol{X}, \boldsymbol{u}\rangle\|_{\psi_2} \leq L\|\boldsymbol{u}\|_2$ hold for any $\boldsymbol{u} \in \mathbb{R}^n$. A random vector $\boldsymbol{X}$ with independent subgaussian entries $X_1, \ldots, X_n$ is a subgaussian vector because [33]

$$\begin{aligned}\|\langle \boldsymbol{X}, \boldsymbol{u}\rangle\|_{\psi_2} &\leq c\sqrt{\sum_{i=1}^n \boldsymbol{u}_i^2 \|X_i\|_{\psi_2}^2} \\ &\leq c \max_{1\leq i \leq n}\|X_i\|_{\psi_2}\|\boldsymbol{u}\|.\end{aligned} \tag{30}$$

Clearly, if in addition $\{X_i\}$ are centered and has unit variance, then $\boldsymbol{X}$ is also isotropic. In particular, the standard Gaussian vector on $\mathbb{R}^n$ and the sign vector with *i.i.d.* $1/2$ Bernoulli entries are isotropic and subgaussian. Isotropic and subgaussian random vectors also include the vectors with the normalized volume measure on various convex symmetric bodies , for example, the unit balls of ℓ_p^n for $2 \leq p \leq \infty$ [29].

We reformulate the ℓ_1−CMSV for sensing matrices with subgaussian entries using empirical processes. Suppose the entries of A are *i.i.d.* with subgaussian tails such that $\mathbb{E}\|A\boldsymbol{u}\|_2^2 = m\|\boldsymbol{u}\|_2^2$ for any $\boldsymbol{u} \in \mathbb{R}^n$. The rows of A are denoted by $\{\boldsymbol{a}_i^T, i = 1, \ldots, m\}$. Denote $\mathcal{H}_s^n = \{\boldsymbol{u} \in \mathbb{R}^n : \|\boldsymbol{u}\|_2^2 = 1, \|\boldsymbol{u}\|_1^2 \leq s\}$. We note that $\rho_s(A/\sqrt{m}) > 1 - \epsilon$ is a consequence of

$$\begin{aligned}&\sup_{\boldsymbol{u}\in\mathcal{H}_s^n}\left|\frac{1}{m}\boldsymbol{u}^T A^T A\boldsymbol{u} - 1\right| \\ =\; &\sup_{\boldsymbol{u}\in\mathcal{H}_s^n}\left|\frac{1}{m}\sum_{i=1}^m \langle \boldsymbol{a}_i, \boldsymbol{u}\rangle^2 - 1\right| \leq \epsilon.\end{aligned} \tag{31}$$

Define a class of functions parameterized by $\boldsymbol{u}$ as $\mathcal{F}_s \stackrel{\text{def}}{=} \{f_{\boldsymbol{u}}(\cdot) = \langle \boldsymbol{u}, \cdot\rangle : \boldsymbol{u} \in \mathcal{H}_s^n\}$ and denote P_m the empirical measure that puts equal mass at each of the m random variables (observations) $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$, *i.e.,*

$$P_m(\cdot) = \frac{1}{m}\sum_{i=1}^m \delta_{\boldsymbol{a}_i}(\cdot) \tag{32}$$

with $\delta_{\boldsymbol{x}}(\cdot)$ the dirac measure that puts unit mass at $\boldsymbol{x}$. We realize that $\{\frac{1}{m}\sum_{i=1}^{m}\langle \boldsymbol{a}_i, \boldsymbol{u}\rangle^2\}$ is the empirical process $\{P_m(f^2)\}_{f\in\mathcal{F}_s}$. We slightly abuse notation and use $\mathbb{E}f^2$ to denote $\mathbb{E}f^2(\boldsymbol{a})$. Then, our goal is to estimate

$$\mathbb{E}\sup_{f\in\mathcal{F}_s}\left|P_m(f^2)-\mathbb{E}f^2\right| \tag{33}$$

and

$$\mathbb{P}\left\{\sup_{f\in\mathcal{F}_s}\left|P_m(f^2)-\mathbb{E}f^2\right|\right\}, \tag{34}$$

a central topic of the study of empirical processes.

A key concept in studying general Gaussian processes as well as the empirical process $\{P_m(f^2)\}_{f\in\mathcal{F}_s}$ is the γ_p function we are going to define. We need some setup first. For any set $\mathcal{X}$, an *admissible sequence* is a sequence of increasing partitions $\{\mathcal{Q}_k\}_{k\geq 0}$ of $\mathcal{X}$ such that $\text{card}(\mathcal{Q}_0) = 1$ and $\text{card}(\mathcal{Q}_k) = 2^{2^k}$ for $k \geq 1$. By a sequence of increasing partitions, we mean that every set in $\mathcal{Q}_{k+1}$ is contained in some set of $\mathcal{Q}_k$. We will use $Q_k(\boldsymbol{x})$ to denote the unique set in partition $\mathcal{Q}_k$ that contains $\boldsymbol{x}$. The diameter of $Q_k(\boldsymbol{x})$ is denoted by $\Delta(Q_k(\boldsymbol{x}))$. Then we have the following definition for γ_p function associated with a metric space:

Definition 4 [34] *Suppose $(\mathcal{X}, d)$ is a metric space and $p > 0$. We define*

$$\gamma_p(\mathcal{X}, d) = \inf \sup_{\boldsymbol{x}\in\mathcal{X}}\sum_{k\geq 0}2^{k/p}\Delta(Q_k(\boldsymbol{x})), \tag{35}$$

where the infimum is taken over all admissible sequences.

The importance of the γ_p lies in its relationship with the behavior of Gaussian process indexed by a metric space when the metric is induced by the Gaussian process. More precise, suppose $\{X_{\boldsymbol{x}}\}_{\boldsymbol{x}\in\mathcal{X}}$ is a Gaussian process indexed by the metric space $(\mathcal{X}, d)$ with

$$d(\boldsymbol{x}, \boldsymbol{y}) = (\mathbb{E}(X_{\boldsymbol{x}} - X_{\boldsymbol{y}})^2)^{1/2}, \tag{36}$$

then we have

$$c\gamma_2(\mathcal{X}, d) \leq \mathbb{E}\sup_{\boldsymbol{x}\in\mathcal{X}}X_{\boldsymbol{x}} \leq C\gamma_2(\mathcal{X}, d) \tag{37}$$

for some numerical constants c and C. The upper bound (the generic chaining bound) was first established by Fernique [35] and the lower bound is obtained by Talagrand using majorizing measures [36]. The rather difficult concept of majorizing measures has been considerably simplified through the notion of "generic chaining", an idea that dates back to Kolmogorov and is greatly advanced in recently years by Talagrand [34].

With these preparations, we present the major result of [29]:

Theorem 2 [29] *Let $\{\boldsymbol{a}, \boldsymbol{a}_i, i = 1, \ldots, m\} \subset \mathbb{R}^n$ be i.i.d. random vectors which induce a measure μ on $\mathbb{R}^n$, and $\mathcal{F}$ be a subset of the unit sphere of $L_2(\mathbb{R}^n, \mu)$ with $\text{diam}(\mathcal{F}, \|\cdot\|_{\psi_2}) = \alpha$. Then there exist absolute constants c_1, c_2, c_3 such that for any $\epsilon > 0$ and $m \geq 1$ satisfying*

$$m \geq c_1\frac{\alpha^2\gamma_2^2(\mathcal{F}, \|\cdot\|_{\psi_2})}{\epsilon^2}, \tag{38}$$

with probability at least $1 - \exp(-c_2\epsilon^2 m/\alpha^4)$,

$$\sup_{f\in\mathcal{F}}\left|\frac{1}{m}\sum_{k=1}^{m}f^2(\boldsymbol{a}_k)-\mathbb{E}f^2(\boldsymbol{a})\right| \leq \epsilon. \tag{39}$$

Furthermore, if $\mathcal{F}$ is symmetric, we have

$$\mathbb{E}\sup_{f\in\mathcal{F}}\left|\frac{1}{m}\sum_{k=1}^{m}f^2(\boldsymbol{a}_k)-\mathbb{E}f^2(\boldsymbol{a})\right|$$
$$\leq c_3\max\left\{\alpha\frac{\gamma_2(\mathcal{F}, \|\cdot\|_{\psi_2})}{\sqrt{m}}, \frac{\gamma_2^2(\mathcal{F}, \|\cdot\|_{\psi_2})}{m}\right\}. \tag{40}$$

We apply Theorem 2 to estimate the ℓ_1-CMSV. Consider the function set $\mathcal{F} = \mathcal{F}_s = \{f_{\boldsymbol{u}}(\cdot) = \langle \boldsymbol{u}, \cdot\rangle : \|\boldsymbol{u}\|_2^2 = 1, \|\boldsymbol{u}\|_1^2 \leq s\}$. Assume $\boldsymbol{a} \in \mathbb{R}^n$ is isotropic and subgaussian. As a consequence of the isotropy of $\boldsymbol{a}$ and $\|\boldsymbol{u}\|_2 = 1$, we get $\mathcal{F}_s$ is a subset of the unit sphere of $L_2(\mathbb{R}^n, \mu)$. The symmetry of $\mathcal{F}_s$ yields

$$\begin{aligned} \alpha &= \text{diam}(\mathcal{F}_s, \|\cdot\|_{\psi_2}) \\ &= 2\sup_{\boldsymbol{u}\in\mathcal{H}_s^n}\mathbb{E}\langle \boldsymbol{u}, \boldsymbol{a}\rangle^2 = 2. \end{aligned} \tag{41}$$

Now the key is to compute $\gamma_2(\mathcal{F}_s, \|\cdot\|_{\psi_2})$. Due to (37), the problem reduces to the computation of $\mathbb{E}\sup_{\boldsymbol{u}\in\mathcal{H}_s^n}X_{\boldsymbol{u}}$ (actually an upper bound suffices), where $\{X_{\boldsymbol{u}}\}_{\boldsymbol{u}\in\mathcal{H}_s^n}$ is the canonical Gaussian process:

$$X_{\boldsymbol{u}} = \langle \boldsymbol{g}, \boldsymbol{u}\rangle, \ \boldsymbol{g} \sim \mathcal{N}(0, \text{I}_n), \boldsymbol{u} \in \mathcal{H}_s^n. \tag{42}$$

Clearly, we have

$$\begin{aligned} \gamma_2(\mathcal{F}_s, \|\cdot\|_{\psi_2}) &\leq c\, \mathbb{E}\sup_{\boldsymbol{u}\in\mathcal{H}_s^n}\langle \boldsymbol{g}, \boldsymbol{u}\rangle \\ &\leq c\, \mathbb{E}\|\boldsymbol{u}\|_1\|\boldsymbol{g}\|_{\infty} \\ &\leq c\, \sqrt{s\log n}. \end{aligned} \tag{43}$$

As a consequence, we have the following theorem:

Theorem 3 *Let the rows of the sensing matrix A be i.i.d. subgaussian and isotropic random vectors. Then there exists constants c_1, c_2, c_3 such that for any $\epsilon > 0$ and $m \geq 1$ satisfying*

$$m \geq c_1\frac{s\log n}{\epsilon^2}, \tag{44}$$

we have

$$\mathbb{E}|1 - \rho_s(A)| \leq c_2\epsilon, \tag{45}$$

and

$$\mathbb{P}\{1 - \epsilon \leq \rho_s(A) \leq 1 + \epsilon\} \geq 1 - \exp(-c_3\epsilon^2 m). \tag{46}$$

This theorem says that at least for subgaussian ensembles (including the Gaussian ensemble and the Bernoulli ensemble), the ℓ_1-CMSV bounds are as tight as the RIC bounds.

V. COMPUTATION OF THE ℓ_1-CMSVS

In this section, we first describe two algorithms to compute a lower bound on the maximal k such that the sufficient condition (15) is satisfied. This gives a way to verify that the ℓ_1 recovery is exact in the noiseless setting. The second part of this section is devoted to the computation of ℓ_1-CMSV and its lower bound.

A. Verifying the Sufficient Condition for Exact ℓ_1 Recovery

Using the ℓ_1-sparsity level to verify the sufficient condition (15) (refer to Section III) is formulated as the following optimization problem:

$$\min_{z:Az=0} \frac{1}{4}\frac{\|z\|_1^2}{\|z\|_2^2}, \tag{47}$$

or equivalently,

$$\max_{z} 4\|z\|_2^2 \ \text{s.t.} \ Az = 0, \|z\|_1 \le 1. \tag{48}$$

Unfortunately, this later optimization problem, which maximizes the ℓ_2 norm over a polyhedron, is NP-hard [37]. By defining $Z = zz^T$ and dropping the rank constraint, we instead use the following semidefinite relaxation to produce a lower bound:

$$(\mathrm{L}_2): \quad \max_{Z:Z\succeq 0} 4\mathrm{trace}(Z)$$
$$\text{s.t. } \mathrm{trace}(AZA^T) = 0, \|Z\|_1 \le 1, \tag{49}$$

where $\|Z\|_1$ is the sum of absolute values of all elements in Z.

Another relaxation based on the ℓ_∞ norm is to solve the following optimization problem (refer to (19)):

$$(\mathrm{L}_\infty): \quad \max_{z} 2\|z\|_\infty \ \text{s.t.} \ Az = 0, \|z\|_1 \le 1, \tag{50}$$

which is solved by the following n linear programs:

$$\max_{z} 2z_i \ \text{s.t.} \ Az = 0, \|z\|_1 \le 1, i = 1, \ldots, n. \tag{51}$$

We observe this linear program subproblem is actually the dual of the linear program subproblem used in [21] to compute $\alpha_1(A, \beta)$ when $\beta = \infty$. In this paper, the linear program subproblems are implemented using the primal-dual algorithm detailed in Chapter 11 of [38]. This algorithm produces results comparable to those in [21] but is significantly faster.

B. Computing ℓ_1-CMSV

An advantage of using the ℓ_1-CMSV as a measure of the "goodness" of a sensing matrix is the *relative* ease of its computation. The computation of ℓ_1-CMSV is equivalent to

$$\min_{x\in\mathbb{R}^n} \|Ax\|_2 \ \text{s.t.} \ \|x\|_1 \le \sqrt{s}, \|x\|_2 = 1. \tag{52}$$

Unfortunately, the above optimization is not convex because of the ℓ_2 constraint $\|x\|_2 = 1$. However, many tools at our disposal can deal with the continuous problem (52), for example, the Lagrange multiplier or the Karush-Kuhn-Tucker condition [39]. We will present an interior point algorithm to directly compute an approximate numerical solution of (52). Since the optimization problem (52) is not convex, there is no guarantee that the solution of the algorithm are the true minima. Thus, we will also present a convex program to compute a lower bound on ℓ_1-CMSV.

The interior point (IP) method provides a general approach to efficiently solve the following general constrained optimization problem:

$$\min_{z\in\mathbb{R}^n} F(z) \ \text{s.t.} \ f(z) \le 0, \ g(z) = 0. \tag{53}$$

The basic idea is to construct and solve a sequence of penalized optimization problems with equality constraints:

$$\min_{z,\sigma} F(z) - \mu\sum_i \log(\sigma_i)$$
$$\text{s.t. } f(z) + \sigma = 0, \ g(z) = 0. \tag{54}$$

By using either a Newton step, which tries to solve the Karush-Kuhn-Tucker equations [39], or a conjugate gradient step using trust regions to solve the penalized problem (54) in each iteration, the interior point approach efficiently generates a sequence of solutions that converge to the solution of (53). Refer to [40]–[42] for more information on this approach.

However, the interior point approach assumes that the objective and constraint functions have continuous second order derivatives, which is not satisfied by the constraint $\|z\|_1 - \sqrt{s} \le 0$. We address the non-differentiability of $f(z) = \|z\|_1 - \sqrt{s}$ by defining $z = z^+ - z^-$ with $z^+ = \max(z, 0) \ge 0$ and $z^- = \max(-z, 0) \ge 0$, which leads to the following augmented optimization:

$$\text{IP:} \quad \min_{z^+,z^-\in\mathbb{R}^n} (z^+ - z^-)^T A^T A(z^+ - z^-)$$
$$\text{subject to } \sum_i z_i^+ + \sum_i z_i^- - s \le 0,$$
$$(z^+ - z^-)^T(z^+ - z^-) = 1,$$
$$z^+ \ge 0, \ z^- \ge 0. \tag{55}$$

This algorithm is employed in [16] to design the transmitting waveform of an OFDM radar for optimal detection and estimation performance.

We briefly describe a semidefinite relaxation (SDR) approach to compute a lower bound on ℓ_1-CMSV. A similar method was employed in [19] to compute an upper bound on sparse variance maximization using the *lifting procedure* for semidefinite programming [43]–[45]. Defining $Z = zz^T$ and dropping the rank constraint transform problem (52) into the following form:

$$\text{SDR:} \quad \min_{Z\succeq 0} \quad \mathrm{trace}(A^T AZ)$$
$$\text{s.t.} \quad \|Z\|_1 \le s, \mathrm{trace}(Z) = 1. \tag{56}$$

Now SDR is a semidefinite programming problem. For a small size problem, a global minimum can be achieved at high precision using SEDUMI [46], SDPT3 [47] or CVX [48]. However, for relatively large n, the interior point algorithm makes the memory requirement prohibitive (see [19] for more discussion). In this paper, we do not consider more efficient implementations of the SDR.

VI. NUMERICAL SIMULATIONS

We use numerical simulations to assess the effectiveness and efficiency of the algorithms presented in Section V. Except for the JN algorithm in Table IV, all other experiments were performed on a platform with a Pentium D CPU@3.40GHz, 2GB RAM, and a Windows XP operating system.

A. Verification of Sufficient Conditions

We first examine the L_2 (49) and L_∞ (50) algorithms for verifying the sufficient condition (15). These two algorithms are compared with the two algorithms d'AE and JN proposed in [20] and [21], respectively. We name the two algorithms d'AE and JN using the abbreviations of the authors' names. Recall that k^* is defined as the maximal k such that (15) is satisfied. In Table I, we show the lower bounds on k^* for a small size Bernoulli matrix with $n = 40$ computed by L_2, L_∞, d'AE and JN. The algorithms of d'AE and JN are provided by the authors for free download online.

TABLE I: Comparison of different verification algorithms for a Bernoulli matrix with leading dimension $n = 40$.

m	lower bounds on k^*				CPU time (s)			
	L_1	L_∞	d'AE	JN	L_1	L_∞	d'AE	JN
20	1	1	2	1	14.29	0.41	1040.40	4.59
24	1	2	2	2	16.11	0.38	694.20	0.69
28	2	3	3	3	15.12	0.37	710.90	16.20
32	2	3	4	3	15.43	0.37	894.08	2.45

TABLE II: Comparison of L_∞ and JN for a Hadamard matrix with leading dimension $n = 256$.

m	lower bounds on k^*		CPU time (s)	
	L_∞	JN	L_∞	JN
25	1	1	3	35
51	2	2	6	70
76	3	3	7	102
102	4	4	9	303
128	5	5	9	544
153	7	7	13	310
179	9	9	15	528
204	12	12	18	1333
230	19	18	18	435

TABLE III: Comparison of L_∞ and JN for a Gaussian matrix with leading dimension $n = 256$.

m	lower bnd on k^*		CPU time (s)	
	L_∞	JN	L_∞	JN
25	1	1	6	91
51	2	2	8	191
76	3	3	10	856
102	4	4	13	5630
128	4	5	16	5711
153	6	6	20	1381
179	7	7	24	3356
204	10	10	25	10039
230	13	14	28	8332

In the next set of experiments, we compare lower bounds on k^* computed by L_∞ and JN, respectively, for $n = 256$. In this case, both the semidefinite relaxation in this paper and that in [20] are too time and memory consuming to compute. The lower bounds and execution times are shown in Table II and III for a Hadamard and a Gaussian matrix, respectively.

Table IV shows the results of L_∞ and JN for a Hadamard matrix with leading dimension $n = 1024$. Note the lower

TABLE IV: Comparison of L_∞ and JN for Gaussian and Hadamard matrices with leading dimension $n = 1024$. In the column head, "G" represents Gaussian matrix and "H" represents Hadamard matrix.

m	lower bounds on k^*			CPU time (s)		
	L_∞(H)	L_∞(G)	JN(G)	L_∞(H)	L_∞(G)	JN(G)
102	3	2	2	182	136	457
204	4	4	4	501	281	1179
307	6	6	6	872	510	2235
409	8	7	7	1413	793	3659
512	11	10	10	1914	990	5348
614	14	12	12	1362	1309	7156
716	18	15	15	1687	1679	9446
819	24	20	21	1972	2033	12435
921	37	29	32	2307	2312	13564

bounds computed by JN and the CPU times of JN in table IV are extracted from [21]. We were not able to carry out the computation of JN within reasonable time in our platform.

From Table I, we see that for $n = 40$, d'AE performs the best, and L_∞ and JN give exactly the same results. However, d'AE is very slow in general. For example, even with a first order implementation, the d'AE takes more than 37 hours for matrices of size 350×500, while the L_∞ takes less than 3 minutes to finish the computation. From Table II, III and IV, we see that L_∞ and JN produces comparable results. However, our L_∞ algorithm performs much faster than the JN algorithm. Because the two algorithms solve n linear program subproblems that are dual to each other, they should yield exactly the same results. However, we observe that sometimes the upper bound and lower bound on *the lower bound of k^** computed by JN does not coincide. The difference in speed might come from the implementation. Our implementation of the linear sub-program employs the primal-dual approach detailed in [38] while [21] uses the commercial LP solver `mosekopt` [49].

B. Computation of ℓ_1-CMSV

We next report the test results of the IP and SDR algorithms for computing the ℓ_1-CMSV and its lower bound, respectively. The interior point algorithms IP is implemented using the MATLAB® function *fmincon*. The SDR is solved using CVX [48] with the default SDPT3 solver.

We fist test IP and SDR on a Gaussian matrix $A \in \mathbb{R}^{20 \times 60}$ for $s = 5$. Due to the existence of local minima, we need to run IP several times and select the minimal function value among all the trials as the ℓ_1-CMSV. Fifty random initial points on the unit sphere in $\mathbb{R}^{60}$ are generated for IP. The SDR only runs once. The results are shown in Table V. In this example, SDR over-relaxes the problem and produces a zero ℓ_1-CMSV.

TABLE V: Comparison of IP and SDR for a Gaussian matrix.

	$\min F(\boldsymbol{z}^*)$	$\mathrm{mean} F(\boldsymbol{z}^*)$	$\mathrm{std} F(\boldsymbol{z}^*)$	mean time (s)
IP	0.0666	0.7133	0.3661	2.8903
SDR	0.0000	0.0000	N/A	53.1583

TABLE VI: IP for a Bernoulli matrix $A \in \mathbb{R}^{50 \times 500}$.

	$\min F(\boldsymbol{z}^*)$	$\mathrm{mean} F(\boldsymbol{z}^*)$	$\mathrm{std} F(\boldsymbol{z}^*)$	mean time (s)
IP	0.000371	0.007472	0.004545	123.6456
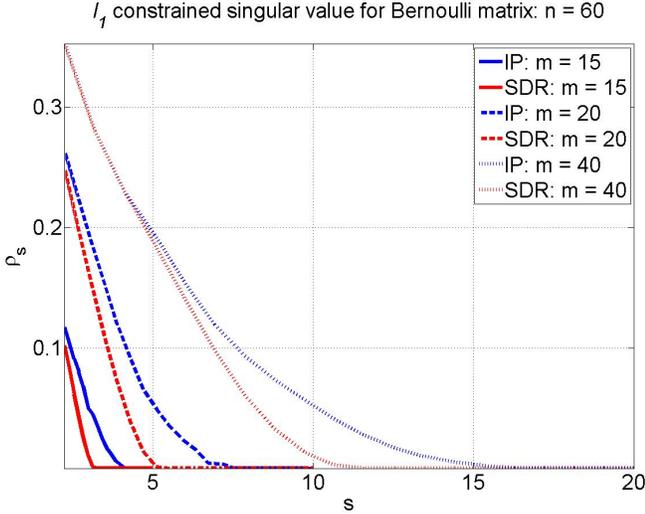
Fig. 1: ℓ_1-CMSV ρ_s and its bound as a function of s for Bernoulli matrix with $n = 60$ and $m = 10, 20, 40$.

The IP is also tested for a Bernoulli matrix $A \in \mathbb{R}^{50 \times 500}$ with results shown in Table VI. The CVX implementation of SDR takes too much memory to run for $n = 500$.

We compare the ℓ_1-CMSVs ρ_s and their bounds as a function of s computed by IP and SDR, respectively, for Bernoulli random matrices. We consider a small-scale problem with $n = 60$ and $m = 10, 20, 40$. A matrix $B \in \mathbb{R}^{40 \times 60}$ with entries $\{+1, -1\}$ following $\frac{1}{2}$ Bernoulli distribution is generated. For $m = 10, 20, 40$, the corresponding Bernoulli matrix A is obtained by taking the first m rows of B. The columns of A are then normalized to have unit norm. The normalization implies that $\rho_s \leq \rho_1 = 1$. The IP uses 30 random initial points. As illustrated in Figure 1, the ℓ_1-CMSVs and their bounds decrease very fast as s increases. For fixed s, increasing m generally (but not necessarily, as shown in Figure 2) increases the ℓ_1-CMSV and their bounds.

In Figure 2, the ℓ_1-CMSV ρ_s is plotted as a function of m with varying parameter values: $s = 5, 10$ and 20. With s fixed, the two algorithms (IP and SDR) are run for $A \in \mathbb{R}^{m \times n}$, with m increasing from $2s$ to $n = 60$. For each m, the construction of A follows the procedure described in the previous paragraph. The discrete nature of adding rows to A while increasing m makes the curves in Figure 2 not as smooth as those in Figure 1. The ρ_s increases with m in general, but local decreases do happen. The gap between values computed by IP and SDR is also clearly seen for medium s.

VII. CONCLUSIONS

In this paper, a new measure of a sensing matrix's incoherence, the ℓ_1-CMSV, is proposed to quantify the stability of sparse signal reconstruction. It is demonstrated that the
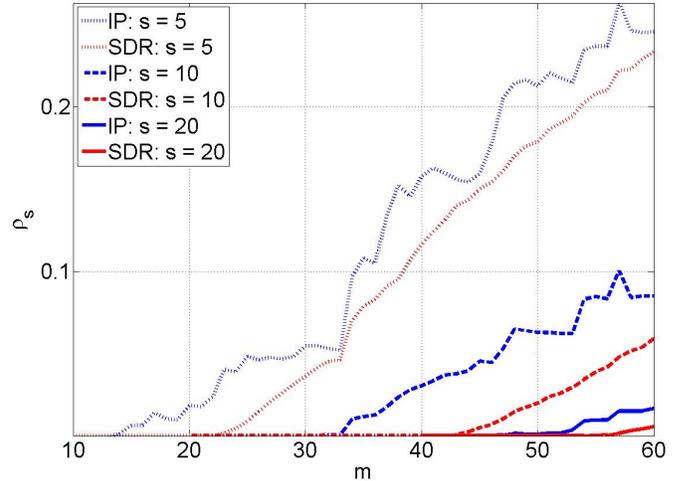
Fig. 2: ℓ_1-CMSV ρ_s and its bound as a function of m for Bernoulli matrix with $n = 60$ and $s = 5, 10, 20$.

reconstruction errors of the Basis Pursuit, the Dantzig selector, and the LASSO estimator are concisely bounded using the ℓ_1-CMSV. A generic chaining argument shows that the ℓ_1-CMSV is bounded away from zero with high probability for the subgaussian ensemble, as long as the number of measurements is relatively large. One interior point program and one semidefinite program are presented to compute the ℓ_1-CMSV and its lower bound, respectively. Numerical simulations assess the algorithms' performance. The ℓ_1-CMSV provides a computationally amenable measure of incoherence that can be used for optimal design.

As a by product, two algorithms are designed to verify the sufficient conditions guaranteeing the uniqueness of ℓ_1-based recovery. The ℓ_∞ relaxation based algorithm is shown to produce comparable results with the state-of-the-art algorithms, and performs much faster.

APPENDIX A
PROOF OF THEOREM 1

In this appendix, we derive the error bounds presented in Theorem 1.

Proof of Theorem 1: We strictly follow the two-step procedure expounded in Section III.

1) In order to establish the ℓ_1-sparsity of the error vector in the first step, we suppose $S = \mathrm{supp}(\boldsymbol{x})$ and $|S| = \|\boldsymbol{x}\|_0 = k$. Define the error vector $\boldsymbol{h} = \hat{\boldsymbol{x}} - \boldsymbol{x}$. For any vector $\boldsymbol{z} \in \mathbb{R}^n$ and any index set $S \subseteq \{1, \ldots, n\}$, we use $\boldsymbol{z}_S \in \mathbb{R}^{|S|}$ to represent the vector whose elements are those of $\boldsymbol{z}$ indicated by S.

We first deal with the BP and the DS. As observed by Candés in [17], the fact that $\|\hat{\boldsymbol{x}}\|_1 = \|\boldsymbol{x} + \boldsymbol{h}\|_1$ is the minimum among all $\boldsymbol{z}$ satisfying the constraints in (3) and (4), *together with* the fact that the true signal $\boldsymbol{x}$ satisfies the constraints as required by the conditions imposed on the noise in Theorem

1, imply that $\|\boldsymbol{h}_{S^c}\|_1$ cannot be very large. To see this, we observe that

$$
\begin{aligned}
\|\boldsymbol{x}\|_1 &\geq \|\boldsymbol{x}+\boldsymbol{h}\|_1 \\
&= \sum_{i\in S}|\boldsymbol{x}_i+\boldsymbol{h}_i| + \sum_{i\in S^c}|\boldsymbol{x}_i+\boldsymbol{h}_i| \\
&\geq \|\boldsymbol{x}_S\|_1 - \|\boldsymbol{h}_S\|_1 + \|\boldsymbol{h}_{S^c}\|_1 \\
&= \|\boldsymbol{x}\|_1 - \|\boldsymbol{h}_S\|_1 + \|\boldsymbol{h}_{S^c}\|_1. \quad (57)
\end{aligned}
$$

Therefore, we obtain $\|\boldsymbol{h}_{S^c}\|_1 \leq \|\boldsymbol{h}_S\|_1$, which leads to

$$
\begin{aligned}
\|\boldsymbol{h}\|_1 &= \|\boldsymbol{h}_S\|_1 + \|\boldsymbol{h}_{S^c}\|_1 \\
&= 2\|\boldsymbol{h}_S\|_1 \\
&\leq 2\sqrt{k}\|\boldsymbol{h}_S\|_2 \\
&\leq 2\sqrt{k}\|\boldsymbol{h}\|_2, \quad (58)
\end{aligned}
$$

where for the next to the last inequality we used the Cauchy-Schwart inequality. Inequality (58) is equivalent to

$$
s(\boldsymbol{h}) \leq 4k. \quad (59)
$$

We now continue to establish the ℓ_1 sparsity of the error vector for the LASSO (5). We borrow ideas from [50] (see also [25]). Since the noise $\boldsymbol{w}$ satisfies $\|A^T\boldsymbol{w}\|_\infty \leq \kappa\lambda_n\sigma$ for some small $\kappa > 0$ and $\hat{\boldsymbol{x}}$ is a solution to (5), we have

$$
\frac{1}{2}\|A\hat{\boldsymbol{x}}-\boldsymbol{y}\|_2^2 + \lambda_n\sigma\|\hat{\boldsymbol{x}}\|_1 \leq \frac{1}{2}\|A\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \lambda_n\sigma\|\boldsymbol{x}\|_1.
$$

Consequently, substituting $\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{w}$ yields

$$
\begin{aligned}
\lambda_n\sigma\|\hat{\boldsymbol{x}}\|_1 &\leq \frac{1}{2}\|A\boldsymbol{x}-\boldsymbol{y}\|_2^2 - \frac{1}{2}\|A\hat{\boldsymbol{x}}-\boldsymbol{y}\|_2^2 + \lambda_n\sigma\|\boldsymbol{x}\|_1 \\
&= \frac{1}{2}\|\boldsymbol{w}\|_2^2 - \frac{1}{2}\|A(\hat{\boldsymbol{x}}-\boldsymbol{x})-\boldsymbol{w}\|_2^2 + \lambda_n\sigma\|\boldsymbol{x}\|_1 \\
&= \frac{1}{2}\|\boldsymbol{w}\|_2^2 - \frac{1}{2}\|A(\hat{\boldsymbol{x}}-\boldsymbol{x})\|_2^2 \\
&\quad + \langle A(\hat{\boldsymbol{x}}-\boldsymbol{x}),\boldsymbol{w}\rangle - \frac{1}{2}\|\boldsymbol{w}\|_2^2 + \lambda_n\sigma\|\boldsymbol{x}\|_1 \\
&\leq \langle A(\hat{\boldsymbol{x}}-\boldsymbol{x}),\boldsymbol{w}\rangle + \lambda_n\sigma\|\boldsymbol{x}\|_1 \\
&= \langle \hat{\boldsymbol{x}}-\boldsymbol{x}, A^T\boldsymbol{w}\rangle + \lambda_n\sigma\|\boldsymbol{x}\|_1.
\end{aligned}
$$

Using the Cauchy-Swcharz type inequality, we get

$$
\begin{aligned}
\lambda_n\sigma\|\hat{\boldsymbol{x}}\|_1 &\leq \|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_1\|A^T\boldsymbol{w}\|_\infty + \lambda_n\sigma\|\boldsymbol{x}\|_1 \\
&= \kappa\lambda_n\sigma\|\boldsymbol{h}\|_1 + \lambda_n\sigma\|\boldsymbol{x}\|_1,
\end{aligned}
$$

which leads to

$$
\|\hat{\boldsymbol{x}}\|_1 \leq \kappa\|\boldsymbol{h}\|_1 + \|\boldsymbol{x}\|_1.
$$

Therefore, similar to the argument in (57) we have

$$
\begin{aligned}
&\|\boldsymbol{x}\|_1 \\
&\geq \|\hat{\boldsymbol{x}}\|_1 - \kappa\|\boldsymbol{h}\|_1 \\
&= \|\boldsymbol{x}+\boldsymbol{h}_{S^c}+\boldsymbol{h}_S\|_1 - \kappa(\|\boldsymbol{h}_{S^c}+\boldsymbol{h}_S\|_1) \\
&\geq \|\boldsymbol{x}+\boldsymbol{h}_{S^c}\|_1 - \|\boldsymbol{h}_S\|_1 - \kappa(\|\boldsymbol{h}_{S^c}\|_1 + \|\boldsymbol{h}_S\|_1) \\
&= \|\boldsymbol{x}\|_1 + (1-\kappa)\|\boldsymbol{h}_{S^c}\|_1 - (1+\kappa)\|\boldsymbol{h}_S\|_1,
\end{aligned}
$$

where $S = \operatorname{supp}(\boldsymbol{x})$. Consequently, we have

$$
\|\boldsymbol{h}_{S^c}\|_1 \leq \frac{1+\kappa}{1-\kappa}\|\boldsymbol{h}_S\|_1.
$$

Therefore, an argument similar to the one leading to (58) yields

$$
\|\boldsymbol{h}\|_1 \leq \frac{2}{1-\kappa}\sqrt{k}\|\boldsymbol{h}\|_2, \quad (60)
$$

or equivalently,

$$
s(\boldsymbol{h}) \leq \frac{4k}{(1-\kappa)^2}. \quad (61)
$$

2) We now turn to obtain an upper bound on $\|A\boldsymbol{h}\|_2$. For the BP (3), this is trivial because both $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ satisfy constraint $\|\boldsymbol{y}-A\boldsymbol{z}\| \leq \epsilon$ in (3). The triangle inequality yields

$$
\begin{aligned}
\|A\boldsymbol{h}\|_2 &= \|A(\hat{\boldsymbol{x}}-\boldsymbol{x})\|_2 \\
&\leq \|A\hat{\boldsymbol{x}}-\boldsymbol{y}\|_2 + \|\boldsymbol{y}-A\boldsymbol{x}\|_2 \\
&\leq 2\epsilon. \quad (62)
\end{aligned}
$$

It then follows from Definition 3 that

$$
\rho_{4k}\|\boldsymbol{h}\|_2 \leq \|A\boldsymbol{h}\|_2 \leq 2\epsilon. \quad (63)
$$

Hence, we get

$$
\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2 \leq \frac{2\epsilon}{\rho_{4k}}. \quad (64)
$$

For the DS (4), as shown in [23], the condition on noise $\|A^T\boldsymbol{w}\|_\infty \leq \lambda_n\sigma$ and the constraint in the Dantzig selector (4) yield

$$
\|A^T A\boldsymbol{h}\|_\infty \leq 2\lambda_n\sigma \quad (65)
$$

because

$$
\begin{aligned}
A_j^T(\boldsymbol{w}-\hat{\boldsymbol{r}}) &= A_j^T[(\boldsymbol{y}-A\boldsymbol{x})-(\boldsymbol{y}-A\hat{\boldsymbol{x}})] \\
&= A_j^T(A\hat{\boldsymbol{x}}-A\boldsymbol{x}) = A_j^T A\boldsymbol{h}, \quad (66)
\end{aligned}
$$

where $\hat{\boldsymbol{r}} = \boldsymbol{y}-A\hat{\boldsymbol{x}}$ is the residual corresponding to the Dantzig selector solution $\hat{\boldsymbol{x}}$. Therefore, we obtain an upper bound on $\|A\boldsymbol{h}\|_2^2$ as follows:

$$
\begin{aligned}
\boldsymbol{h}^T A^T A\boldsymbol{h} &= |\sum_{i=1}^{n}\boldsymbol{h}_i(A^T A\boldsymbol{h})_i| \\
&\leq \sum_{i=1}^{n}|\boldsymbol{h}_i|\cdot|(A^T A\boldsymbol{h})_i| \\
&\leq 2\lambda_n\sigma\|\boldsymbol{h}\|_1. \quad (67)
\end{aligned}
$$

Equation (67), the definition of ρ_{4k}, and equation (58) together yield

$$
\begin{aligned}
\rho_{4k}^2\|\boldsymbol{h}\|_2^2 &\leq \boldsymbol{h}^T A^T A\boldsymbol{h} \\
&\leq 2\lambda_n\sigma\|\boldsymbol{h}\|_1 \\
&\leq 4\lambda_n\sqrt{k}\sigma\|\boldsymbol{h}\|_2. \quad (68)
\end{aligned}
$$

We conclude that

$$
\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2 \leq \frac{4\sqrt{k}}{\rho_{4k}^2}\lambda_n\sigma. \quad (69)
$$

Now we establish an upper bound on $\|A\boldsymbol{h}\|_2^2$ for the LASSO (5) using a procedure similar to the one used for the DS given above. We need to establish a bound on

$$
\begin{aligned}
&\|A^T A\boldsymbol{h}\|_\infty \\
\leq\ &\|A^T(\boldsymbol{y} - A\boldsymbol{x})\|_\infty + \|A^T(\boldsymbol{y} - A\hat{\boldsymbol{x}})\|_\infty \\
\leq\ &\|A^T\boldsymbol{w}\|_\infty + \|A^T(\boldsymbol{y} - A\hat{\boldsymbol{x}})\|_\infty \\
=\ &\kappa\lambda_n\sigma + \|A^T(\boldsymbol{y} - A\hat{\boldsymbol{x}})\|_\infty.
\end{aligned}
\tag{70}
$$

We again follow the procedure in [50] (see also [25]) to estimate $\|A^T(\boldsymbol{y} - A\hat{\boldsymbol{x}})\|_2$. Since $\hat{\boldsymbol{x}}$ is the solution to (5), the optimality condition yields that

$$
A^T(\boldsymbol{y} - A\hat{\boldsymbol{x}}) \in \lambda_n\sigma\partial\|\hat{\boldsymbol{x}}\|_1,
\tag{71}
$$

where $\partial\|\hat{\boldsymbol{x}}\|_1 = [-1, 1]^n$ is the subgradient of $\|\cdot\|_1$ evaluated at $\hat{\boldsymbol{x}}$.

As a consequence, we obtain

$$
\|A^T(\boldsymbol{y} - A\hat{\boldsymbol{x}})\|_\infty \leq \lambda_n\sigma.
\tag{72}
$$

Following the same lines in (67), we get

$$
\|A\boldsymbol{h}\|_2^2 \leq (\kappa + 1)\lambda_n\sigma\|\boldsymbol{h}\|_1.
\tag{73}
$$

Then, Equation (60), (70) and (72)

$$
\begin{aligned}
\rho^2_{\frac{4k}{(1-\kappa)^2}}\|\boldsymbol{h}\|_2^2 &\leq \|A\boldsymbol{h}\|_2^2 \\
&\leq (\kappa + 1)\lambda_n\sigma\frac{\sqrt{4k}}{1-\kappa}\|\boldsymbol{h}\|_2.
\end{aligned}
\tag{74}
$$

As a consequence, we get

$$
\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{1+\kappa}{1-\kappa} \cdot \frac{2\sqrt{k}}{\rho^2_{\frac{4k}{(1-\kappa)^2}}}\lambda_n\sigma.
\tag{75}
$$

∎

References

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[4] E. G. Larsson and Y. Selen, "Linear regression with a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 451–460, Feb. 2007.

[5] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[6] D. Model and M. Zibulevsky, "Signal reconstruction in sensor arrays using sparse representations," *Signal Processing*, vol. 86, pp. 624–638, Mar. 2006.

[7] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM Journal on Applied Mathematics*, vol. 49, no. 3, pp. 906–931, 1989.

[8] M. A. Sheikh, S. Sarvotham, O. Milenkovic, and R. G. Baraniuk, "DNA array decoding from nonlinear measurements by belief propagation," in *Proc. IEEE Workshop Statistical Signal Processing (SSP 2007)*, Madison, WI, Aug. 2007, pp. 215–219.

[9] H. Vikalo, F. Parvaresh, and B. Hassibi, "On recovery of sparse signals in compressed DNA microarrays," in *Proc. Asilomar Conf. Signals, Systems and Computers (ACSSC 2007)*, Pacific Grove, CA, Nov. 2007, pp. 693–697.

[10] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays," *IEEE J. Sel. Topics Signal Processing*, vol. 2, no. 3, pp. 275–285, June 2008.

[11] H. Vikalo, F. Parvaresh, S. Misra, and B. Hassibi, "Sparse measurements, compressed sampling, and DNA microarrays," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP 2008)*, Las Vegas, NV, Apr. 2008, pp. 581–584.

[12] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *IEEE Radar Conference*, Apr. 2007, pp. 128–133.

[13] M. Herman and T. Strohmer, "Compressed sensing radar," in *IEEE Radar Conference*, May 2008, pp. 1–6.

[14] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, June 2009.

[15] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP 2007)*, Honolulu, HI, Apr. 2007, pp. IV–1357–IV–1360.

[16] S. Sen and A. Nehorai, "Sparsity-based multi-target tracking using ofdm radar," Submitted to *IEEE Trans. Signal Process.*

[17] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences, Paris, Serie I*, vol. 346, 2008.

[18] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *J. Amer. Math. Soc.*, vol. 22, pp. 211–231, July 2009.

[19] A. d'Aspremont, L. El Ghaoui, M. Jordan, and G. R. G. Lanckriet, "A direct formulation for sparse PCA using semidefinite programming," *SIAM Review*, vol. 49, no. 3, pp. 434–448, 2007.

[20] A. d'Aspremont and L. El Ghaoui, "Testing the nullspace property using semidefinite programming," *ArXiv e-prints*, Nov. 2010.

[21] A. Juditsky and A. S. Nemirovski, "On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization," *ArXiv e-prints*, May 2010.

[22] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[23] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, pp. 2313–2351, 2007.

[24] R. Tibshirani, "Regression shrinkage and selection via lasso," *J. Roy. Statist. Soc. Ser. B*, vol. 58, pp. 267–288.

[25] P. Bickel, Y. Ritov, and A. Tsybakov, "Simultaneous analysis of Lasso and Dantzig selector," *Annals of Statistics*, vol. 37, no. 4, pp. 1705–1732, 2009.

[26] P. J. Bickel, "Discussion of the Dantzig selector: statistical estimation when p is much larger than n, by e. j. candès and t. tao," *Annals of Stat.*, pp. 2352–2357, 2007.

[27] N. Meinshausen and B. Yu, "Lasso-type recovery of sparse representations for high-dimensional data," *Ann. Statist.*, vol. 37, pp. 246–270, 2009.

[28] R. DeVore R. Baraniuk, M. Davenport and M. B. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[29] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Reconstruction and subgaussian operators in asymptotic geometric analysis," *Geometric And Functional Analysis*, pp. 1248–1282, Nov. 2007.

[30] Y. Zhang, "A simple proof for recoverability of ℓ_1-minimization: go over or under?," Tech. Rep., Rice CAAM Department, 2005.

[31] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[32] D. Donoho, "High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension," Technical report, Department of Statistics, Stanford University, 2004.

[33] S. Mendelson and N. Tomczak-Jaegermann, "A subgaussian embedding theorem," *Israel Journal of Mathematics*, pp. 349–364, Mar. 2008.

[34] M. Talagrand, *The generic chaining: upper and lower bounds of stochastic processes*, Springer, 2005.

[35] X. Fernique, *Régularité des trajectoires des fonctiones aléatoires gaussiennes*, Ecole d'Eté de Probabilités de St-Flour 1974, Lecture Notes in Mathematics 480, Springer-Verlag, 1975.

[36] M. Talagrand, "Regularity of Gaussian processes," *Acta Math.*, vol. 159, pp. 99–149, 1987.

[37] H. Bodlaender, P. Gritzmann, V. Klee, and J. Leeuwen, "Computational complexity of norm-maximization," *Combinatorica*, vol. 10, pp. 203–225, 1990.

[38] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[39] A. Sofer S. G. Nash, *Linear and nonlinear programming*, McGraw-Hill, New York, NY, 1996.

[40] R. H. Byrd, J. C. Gilbert, and J. Nocedal, "A trust region method based on interior point techniques for nonlinear programming," *Mathematical Programming*, vol. 89, no. 1, pp. 149–185, 2000.

[41] R. H. Byrd, Mary E. Hribar, and Jorge Nocedal, "An interior point algorithm for large-scale nonlinear programming," *SIAM Journal on Optimization*, vol. 9, no. 4, pp. 877–900, 1999.

[42] R. A. Waltz, J. L. Morales, J. Nocedal, and D. Orban, "An interior algorithm for nonlinear optimization that combines line search and trust region steps," *Mathematical Programming*, vol. 107, no. 3, pp. 391–408, 2006.

[43] L. Lovász and A. Schrijver, "Cones of matrices and set-functions and 0-1 optimization," *SIAM Journal on Optimization*, vol. 1, pp. 166–190, 1991.

[44] F. Alizadeh, "Interior point methods in semidefinite programming with applications to combinatorial optimization," *SIAM Journal on Optimization*, vol. 5, pp. 13–51, 1995.

[45] C. Lemaréchal and F. Oustry, "Semidefinite relaxations and Lagrangian duality with application to combinatorial optimization," *INRIA, Rapport de recherche*, vol. 3710, 1999.

[46] J. Sturm, "Using SEDUMI 1.0x, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11, pp. 625–653, 1999.

[47] R. H Tutuncu, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Mathematical Programming Ser. B*, vol. 95, pp. 189–217, 2003.

[48] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming (web page and software)," June 2009.

[49] K. Andersen E. Andersen, "The MOSEK optimization tools mannual, Version 5.0," http://www.mosek.com.

[50] E. J. Candès and Y. Plan, "Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements," Submitted for publication, 2009.